NIH Public Access
Author Manuscript
J Exp Psychol Hum Percept Perform. Author manuscript; available in PMC 2011 August 19.

Published in final edited form as:
J Exp Psychol Hum Percept Perform. 2010 August ; 36(4): 976–993. doi:10.1037/a0017171.

Psychophysical Reverse Correlation with Multiple Response Alternatives

Huanping Dai and
Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona

Christophe Micheyl
Auditory Perception and Cognition Laboratory, Department of Psychology, University of Minnesota, Minnesota

Abstract

Psychophysical reverse-correlation methods such as the "classification image" technique provide a unique tool to uncover the internal representations and decision strategies of individual participants in perceptual tasks. Over the last thirty years, these techniques have gained increasing popularity among both visual and auditory psychophysicists. However, thus far, principled applications of the psychophysical reverse-correlation approach have been almost exclusively limited to two-alternative decision (detection or discrimination) tasks. Whether and how reverse-correlation methods can be applied to uncover perceptual templates and decision strategies in situations involving more than just two response alternatives remains largely unclear. Here, the authors consider the problem of estimating perceptual templates and decision strategies in stimulus identification tasks with multiple response alternatives. They describe a modified correlational approach, which can be used to solve this problem. The approach is evaluated under a variety of simulated conditions, including different ratios of internal-to-external noise, different degrees of correlations between the sensory observations, and various statistical distributions of stimulus perturbations. The results indicate that the proposed approach is reasonably robust, suggesting that it could be used in future empirical studies.

Keywords

psychophysics; reverse correlation; perceptual templates; decision rules

The perception of all but the simplest stimuli usually involves combining multiple sources of sensory information. For instance, the identification of familiar faces requires combining visual cues related to the shape of the eyes, nose, mouth, and other distinguishing features. In order to understand how complex stimuli such as faces, everyday-life objects, or speech signals are successfully identified, it is essential to determine what stimulus features are used by observers (or listeners), and how these various cues are combined. More specifically, it is desirable to know how much different parts, regions, features, or components of a complex stimulus "weight" in determining an observer's or listener's decisions regarding that stimulus. By measuring and inspecting "stimulus-weighting patterns", experimenters can determine what "internal templates" and "decision strategies" mediate an individual's performance in a visual or auditory perception task.

Correspondence concerning this manuscript should be addressed to: Huanping Dai, Department of Speech, Language, and Hearing Sciences, University of Arizona, 1131 E. 2nd Street, Tucson, Arizona, 85721. hdai@email.arizona.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

While traditional psychophysical methods can provide some insight into perceptual templates and decision strategies, more recently developed approaches, inspired by system-identification techniques used in engineering and neurophysiology (e.g., Ringach & Shapley, 2004), appear to be ideally suited for this purpose (Victor, 2005). These "psychophysical reverse correlation" techniques, as they are commonly referred to (Neri, Parker & Blakemore, 1999; Neri & Levi, 2006; Solomon, 2002), belong to the general category of "molecular" psychophysical methods (Green, 1964). Molecular psychophysics differs from traditional, or "molar", psychophysics in that it involves a detailed, trial-by-trial analysis of the relationship between stimuli and responses. The general principle behind psychophysical reverse-correlation techniques is the analysis of the statistical relationships between random perturbations ("noise") applied to different parts of a stimulus across many trials, and the responses given by the observer (or listener) on corresponding trials. The results of this correlational analysis are commonly referred to as "classification images" in the visual perception literature, and as (spectral or temporal) "weighting patterns" in the auditory perception literature. They provide direct insight into the perceptual templates used by participants in visual or auditory perception tasks, and can be used to determine what perceptual decision strategies mediate performance in those tasks.

Psychophysical reverse-correlation methods come in various flavors. The "classification images" technique (Ahumada, 1996; Ahumada & Beard, 1998; Beard & Ahumada, 1999; Eckstein & Ahumada, 2002; Eckstein, Shimozaki, & Abbey, 2002; Murray, Bennett, & Sekuler, 2002, 2005), which actually originated in early auditory psychophysical studies (Ahumada & Lovell, 1971; Ahumada, Marken, & Sandusky, 1975), the "bubbles" technique (Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002), the "correlational" technique (Knill, 1998; Lutfi, 1995; Richards & Zhu, 1994), and the "conditioned-on-a-single stimulus" technique (Berg, 1989; Berg & Green, 1990) are the most commonly used. The former two are more familiar to visual psychophysicists, while the latter two have been used exclusively by auditory psychophysicists. In the "classification images" technique (see Ahumada, 2002 for a review), noise is added to an image, which may or may not contain a signal (for instance, an oriented Gabor patch) that the observer has to detect. The detailed characteristics of the noise samples applied on different trials are recorded, along with the presence or absence of the signal, and the observer's responses. At the end of the experiment, which usually involves a large number of trials (typically, in the thousands), the noise samples are sorted based on the observer's responses and, optionally, on whether the signal was present or absent. Specifically, noise samples applied on trials on which the observer declared that the signal was absent are averaged together, and subtracted from the average of the noise samples applied on trials on which the observer reported that the signal was present. The resulting difference, which is referred to as a "classification image", provides a picture of the observer's representation of the signal. By comparing empirical classification images with theoretical templates derived under different model assumptions, experimenters can determine which model best captures the observer's behavior in the considered task. The classification-image technique has inspired a large number of publications over the past forty years. It has been used, among other things, to study the perception of contrast (Neri & Heeger, 2002; Shimozaki, Eckstein, & Abbey, 2005), orientation (Mareschal, Dakin, & Bex, 2006; Mareschal, Morgan, & Solomon, 2008; Nagai, Bennett, & Sekuler, 2007), color (Brenner, Granzier, & Smeets, 2007; Bouet & Knoblauch, 2004), luminance (Thomas & Knoblauch, 2005), figure-ground segregation (Neri & Levi, 2007) real or illusory contours (Gold, Murray, Bennett, & Sekuler, 2000; Keane, Lu, & Kellman, 2007), crowding (Nandy & Tjan, 2007), visual search (Tavassoli, van der Linde, Bovik, & Cormack, 2007a, 2007b), attention and neglect (Shimozaki, Chen, Abbey, & Eckstein, 2007; Shimozaki, Kingstone, Olk, Stowe, & Eckstein, 2006), amblyopia (Levi & Klein, 2002, 2003; Levi, Klein, & Chen, 2005, 2007, 2008), and perceptual learning effects (Kurki, Hyvarinen, & Laurinen, 2006; Michel & Jacobs, 2008), among others.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Similar to the classification-image technique, the "correlational" technique (Lutfi, 1995; Richards & Zhu, 1994) involves the addition of random perturbations to a stimulus, which the observer or listener must detect or discriminate. The main difference between the two techniques is in the way the decision weights are estimated. In the correlational approach, as the name indicates, the decision weights are estimated based on the measured correlation between stimulus perturbations and the binary-coded responses of the listener (e.g., 0 for "No" or "Signal absent", and 1 for "Yes" or "Signal present"). Intuitively, a relatively high correlation between the perturbations applied to a given stimulus component and the participant's responses indicates that the considered component has a relatively large weight in the participant's decisions; conversely, a correlation coefficient close to zero indicates that the considered component exerts little or no influence on the participant's decisions. More rigorously, Richards and Zhu (1994) and Lutfi (1995) demonstrated that the point biserial correlation coefficients between Gaussian-distributed stimulus perturbations and binary responses in a signal-in-noise detection task were directly proportional to the underlying decision weights. In other words, the correlation coefficients can be used to estimate the underlying decision weights, up to a proportionality constant. While the value of this constant is usually unknown to experimenters, this does not represent a major problem in most applications, because experimenters are typically more interested in *relative* than in *absolute* decision weights. Richards and Zhu's (1994) and Lutfi's (1995) seminal articles provide a firm theoretical foundation for subsequent applications of the correlational technique in auditory perception studies (e.g., Alexander & Lutfi, 2004; Buus, 1999; Buus, Zhang, & Florentine, 1996; Dai, 2000; Dai, Nguyen, & Green, 1996; Doherty & Lutfi, 1996, 1999; Kortekaas, Buus, & Florentine, 2003; Lutfi, 2001; Lutfi & Jesteadt, 2006; Lutfi & Oh, 1997; Lutfi & Wang, 1999; Qian & Richards, 2008; Richards & Tang, 2006; Stellmack & Lutfi, 1996; Stellmack & Viemeister, 2000; Stellmack, Willihnganz, Wightman, & Lutfi, 1997; Wang, Irwin, & Hautus, 2005; Willihnganz, Stellmack, Lutfi, & Wightman, 1997).

Despite superficial differences, the classification-image and correlational techniques are in fact closely related to each other. The point-biserial correlation coefficient, c_{xr}, between a vector $x = [x^1, \ldots, x^t]$ representing the values of a randomly varying stimulus variable across t trials, and the corresponding series of binary values, $r = [r^1, \ldots, r^t]$, can be expressed as a standardized difference between response-conditional means of the stimulus perturbations, as follows (Kornbrot, 2005; Sheskin, 2000):

$$c_{xr} = \frac{m_1 - m_0}{s_x} \sqrt{p_1 \, p_0} \tag{1}$$

In this equation, m_1 and m_0 denote the mean stimulus perturbations (or noise fields) across trials on which the response variable was equal to 1 (e.g., "Yes") or 0 (e.g., "No"), respectively, while p_1 and p_0 denote the corresponding response proportions, out of the t trials. The denominator, s_x, denotes the standard deviation of x. The first term on the right-hand side of Equation 1 is the basis of the formula used for the computation of classification images as a difference between noise fields (e.g., Abbey & Eckstein, 2002; Ahumada, 2002).

While the classification-image and correlational approaches continue to inspire publications in both vision and hearing, the theory behind these techniques was originally developed in the context of decision models involving a choice between two alternatives, and it is not entirely clear whether and how these approaches can be applied in experiments involving more than two possible response alternatives. Certainly, the classification-image method has been the object of many interesting theoretical developments since its inception, including

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

modifications aimed at coping with nonlinearities and/or sources of uncertainty in the decision process (Abbey & Eckstein, 2002; Murray et al., 2002; Tjan & Nandy, 2006). Some of these developments have considered decision models far more sophisticated than the basic single-template model on the basis of which the classification-image technique was initially developed. For instance, to study the influence of uncertainty on classification-images, Tjan and Nandy (2006) considered a decision model involving a comparison between multiple templates, corresponding to multiple possible positions of the stimuli in a two-alternative letter identification task. However, the specific problem of how the classification-image and correlational analyses can be modified to apply in situations involving more than two response alternatives remains without a clear answer. As a result, principled applications of the classification-image and correlational techniques have thus far been limited to Yes-No detection or two-alternative forced-choice discrimination experiments, including, albeit exceptionally, “same-different” experiments (Dai, Versfeld, & Green, 1996; Wang et al., 2005).

Two-alternative tasks do not lend themselves very well to the exploration of decision strategies in, e.g., object, speech, or face recognition studies, where the use of only two stimulus and response categories is both unnatural and impractical. Although there have been a few, very elegant applications of the classification-image approach to the study of perceptual templates for face categorization (Dotsch, Wigboldus, Langner, & van Knippenberg, 2008; Mangini & Biederman, 2004; Martin-Malivel, Mangini, Fagot, & Biederman, 2006; Nestor & Tarr, 2008), the experiments in these studies always involved a discrimination between two alternatives (e.g., male/female, happy/unhappy, Cruise/Travolta). The constraint of using only two response alternatives in studies such as these poses both practical and fundamental problems. From a practical point of view, it forces the experimenter to either choose an arbitrary subset of stimuli, or to test all pairwise combinations of stimulus categories sequentially; clearly, the latter option quickly becomes impractical as the number of stimulus categories increases. More fundamentally, there is no guarantee that observers use the same perceptual strategies when they have to categorize a stimulus as A vs B, A vs C, or B vs C sequentially, as when they have to identify which of the three possible stimuli (A, B, or C) was presented. The obvious solution to this problem is to offer participants more than two response alternatives. To our knowledge, the only published trace of a study using the classification-image approach to measure perceptual templates in a visual identification task with more than two alternatives is an abstract by Watson (1998) in proceedings of the *Investigative Ophthalmology and Visual Science* journal. The abstract shows classification images corresponding to the letters e, i, and z, which an observer had to identify in noise in a three-alternative forced choice task. Unfortunately, the abstract gives very little details about the way in which the data were collected and analyzed. In recent years, auditory researchers have used a variant of the correlational approach in speech-perception experiments where listeners were given a choice between several possible responses (Apoux & Bacon, 2004; Calandruccio & Doherty, 2007, 2008; Doherty & Turner, 1996; Gilbert & Micheyl, 2005; Mehr, Turner, & Parkinson, 2001; Turner et al., 1998). However, the assumptions on which such applications of the correlational analysis are based remain unclear, and the interpretation of the results obtained in these studies is questionable.

Overall, it appears that the lack of a clear theoretical basis for applications of the correlational technique in experiments involving more than just two response alternatives is stifling efforts to understand how observers or listeners identify visual or auditory stimuli. In this note, a generalization of the psychophysical correlational technique to tasks involving more than just two possible response alternatives is described and tested. The outline of the note is as follows. The first section provides a brief reminder about the general decision-theoretic framework and linear decision model for two-alternative tasks, which form the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

basis on which the correlational approach was initially developed. In the second section, the decision model and optimal (likelihood ratio) decision rule in multiple-alternative identification is described. To make things more concrete, a specific example of five-alternative spectral-shape identification task taken from the auditory-perception literature is described; this example is used on multiple occasions in subsequent sections. In the third section, a modified correlational approach for the estimation of decision weights in stimulus identification experiments involving multiple response alternatives is described. Computer simulations are used to show that this approach yields valid estimates of the underlying decision weights in the example five-alternative spectral-shape identification task. In the following two sections, two important specificities of the proposed analysis are discussed: the sorting of trials into stimulus-specific blocks, and the use of response scores rather than responses. Alternative approaches for estimating decision weights in multiple-alternative tasks, including a modified "classification image" analysis, are considered, and compared to that proposed here. The next three sections of the manuscript are devoted to exploring the conditions under which the proposed correlational approach can or cannot be used; the presence internal noise, correlations between the sensory observations, and non-Gaussian perturbations are simulated, and their influence on the measured decision weights analyzed. The final section addresses the issue of how non-linearities in the mapping between stimulus and decision variable can affect experimenters' ability to estimate decision weights using the proposed correlational approach.

General Decision-Theoretic Framework and Linear Decision Model

Before delving into the specifics of the decision model for the *m*-alternative identification task, it may be helpful to present briefly some essential aspects of the general psychophysical decision-theoretic framework and, more specifically, of the linear decision model. In the general decision-theoretic framework, the ideal observer bases its decisions on a variable or "decision statistic", which is monotonically related to the ratio of the likelihoods of the different stimulus alternatives (Green & Swets, 1966; Macmillan & Creelman, 2005; Wickens, 2001). The likelihood ratio is a real-valued function of the observations; thus, the domain of this function is the *n*-dimensional observation space. Consequently, the decision statistic can be represented geometrically as a surface in a "decision space" of dimension n+1, where axes 1 to n correspond to the observations, and axis n+1 is used to represent the value of the decision statistic. While the n+1 axes are orthogonal to each other, the observations need not be independent of each other. The final stage in the decision model consists in using the value of the decision variable to choose a response. In the simplest and most common case, there are only two response alternatives, R_1 and R_2 (for instance: "Yes" and "No", or "1st interval" and "2nd interval"), and the decision rule is to select one response (say, R_1) if the decision variable exceeds some criterion value, k, and the other response (say, R_2) otherwise.

The linear version of the decision model is characterized by the decision statistic, Z, being obtained as a linear function of the observations, $y_1, \ldots y_n$:

$$Z = w_1 y_1 + \ldots + w_n y_n, \tag{2}$$

where the coefficients, $w_1, \ldots, w_n$, represent "weights" assigned to the observations by the observer when making his/her decision. Geometrically, these weights correspond to the slopes of the lines formed by the intersection of the (hyper-)plane decision surface with n two-dimensional planes each being parallel to a plane defined by one of the observation axes (i.e., axes 1 to n), and the decision-statistic axis (i.e., axis n+1).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

If the observations corresponding to the ith dimension of the observation space come from a Gaussian distribution with constant variance, then the decision weight, w_i, can in principle be estimated (up to an unknown proportionality constant[1]) using the point-biserial correlation coefficient, c_i, between the series of observations $y_i=[y_i^1,\ldots,y_i^t]$ across t trials, and the corresponding series of binary responses, $r = [r^1, \ldots, r^t]$. The exact mathematical relationship between the decision weights and the point-biserial correlation coefficients is given in Richards and Zhu (1994). In practice, experimenters usually do not have access to the sensory observations. However, for most basic dimensions of auditory or visual perception, there exists some suitably transformed physical measure of the stimulus (e.g., image contrast in percents, or sound level in decibels) to which the sensory observations are, at least approximately, linearly related. Moreover, the requirement that the Gaussian distributions from which the n observations are drawn all have the same variance can be relaxed if these variances are known or can be estimated, as their influence on the measured correlation coefficients can then be corrected for (Richards & Zhu, 1994).

Decision Model and Optimal Decision Rule for *m*-Alternative Identification

In trying to extend the correlational technique to handle psychophysical tasks that involve more than two stimulus and response alternatives, a first complication relates to the fact that the optimal observer can no longer base its decisions on a single likelihood ratio. In an identification task with m stimulus alternatives, $S_1, \ldots, S_m$, and m corresponding response alternatives, $R_1, \ldots, R_m$, the observer must compare m likelihoods (Green & Birdsall, 1978; Nolte & Jaarsma, 1967; Pelli, Burns, Farell, & Moore-Page, 2006; Swets, Tanner, & Birdsall, 1961; van Trees, 2001). Specifically, denoting as $f(Y \mid S_i)$ the probability density function of the n-dimensional random vector of observations, $Y = [y_1, \ldots, y_n]$, conditioned on stimulus alternative i ($i = 1, \ldots, m$), and using Bayes rule, we have:

$$L_{i,j}(Y) \propto \frac{f(Y|S_i)}{f(Y|S_j)} \tag{3}$$

where $L_{i,j}(Y)$ denotes the likelihood ratio of S_i over S_j; if S_i and S_j are equally likely *a priori*, the proportionality relationship ($\propto$) can be replaced by a strict equality. In this context, a likelihood-ratio optimal decision rule is as follows: given a vector of observations, $Y = [y_1, \ldots, y_n]$, respond R_i (indicating the belief that stimulus alternative S_i was presented) if and only if $L_{i,j}(Y) \geq 1$ for all $j \neq i$. This decision rule is equivalent to choosing the most likely stimulus alternative, given the observations, i.e., choose response i such that $L_{i,j}(Y)$ is maximized. If the m stimulus alternatives are all equally likely *a priori*, this maximum-likelihood (ML) decision rule is equivalent to the maximum *a posteriori* (MAP) decision rule.

To make things more concrete, we take the example of a five-alternative auditory spectral-shape-identification experiment. For readers who are not familiar with the auditory psychophysics literature, we should mention that spectral shape is a critically important attribute of sounds, which plays a key role in both speech and music perception. For instance, vowel sounds differ primarily from each other in their spectral envelopes;

[1]The unknown proportionality constant in the relationship between decision weights and correlation coefficients stems from the fact that the latter depend not just on the underlying weights, but also on the variance of R – for instance, if the signal strength is too high and the observer responds "Yes" on all trials, the correlation coefficient will equal zero, even though the underlying decision weight may be high. One consequence of the unknown proportionality constant is that only *relative* weights, $w_1', \ldots, w_n'$, where $w_k' = w_k / \sum w_k$, and not the *absolute* weights, $w_1, \ldots, w_n$, can be estimated. Fortunately, this limitation is rarely of practical relevance, because in most applications, experimenters are interested only in the relative weights.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

especially, the location of spectral peaks (also known as "formants") along the frequency axis. Therefore, the perception of spectral shape, or "spectral profile analysis" (Green, 1988), is a topic of considerable interest for auditory psychophysicists. Figure 1 illustrates five sound spectra, which are fairly typical of those used in spectral-profile-analysis studies. In fact, these five stimulus spectra were used in an experiment by Dai, Nguyen, and Green (1996). Each stimulus consists of five spectral components, and the stimulus spectra only differ from each other by which of the five components has a higher level than the other four components. The five resulting stimuli, shown in Figure 1, are labeled S_1 to S_5, where the subscript denotes which of the five spectral components (numbered from low to high) receives the level increment. On each trial, one of these five stimuli is selected at random (with equal probability), and presented to the listener. The listener's task is to identify which stimulus was presented or, equivalently, which of the five spectral components had a higher level. To ensure that listeners perform the task by comparing the level across the frequency components, rather than by comparing the level of each component with a fixed reference (in memory), the overall level of the stimuli is roved (i.e., randomly varied) over a wide range across trials (Green, 1988). The magnitude of the level increment, or signal level (abbreviated "DL"), can be thought of in general terms as the "signal strength". In a visual perception experiment, the signal strength may correspond, for instance, to the contrast between the signal and the background in an image. In general, an increase in signal strength is expected to lead to a higher probability of correct response.

In modeling this situation, we assume that the decision maker has access to the individual levels of the components in the stimulus, $l_1,\ldots, l_n$. For real listeners, this presumes that the components are sufficiently separated in frequency, so that they do not appreciably interact within the bandwidths of peripheral auditory filters. Note that we make no assumption regarding the existence of correlations between the stimuli corresponding to the different alternatives. Real-life signals, such as speech tokens, are not in general strictly orthogonal. The analysis presented here does not require that the stimuli be uncorrelated. No special significance should be attached to the fact that in this example, the number of stimulus components, n, equals the number of response alternatives, m; the analysis described below places no restriction on the values of m and n.

Assuming that the sensory observations corresponding to the five stimulus-component levels, $l_1, \ldots, l_n$, are added Gaussian perturbations (which may be of external or internal origin), an optimal decision strategy for the considered m-alternative identification task is derived in Appendix. It consists in selecting response R_k such that the difference between the observation evoked by the kth stimulus component and the mean of the observations evoked by the other four components is largest. The weighting patterns corresponding to this optimal decision strategy consist in assigning a weight of α (where α is an arbitrary positive scalar) to the incremented component, and weights of −α/4 to each of the other components. It is worth stressing that in the m-alternative situation considered here, the optimal decision strategy involves m weighting patterns, and these m "candidate" weighting patterns are all used by the observer on each trial, in order to decide which of the stimuli was most likely presented. This differs from the two-alternative case, where the optimal decision strategy usually involves a single weight vector. Specifically, for the m-alternative identification task considered here, the optimal decision strategy involves computing the inner-product of the incoming vector of spectral-components levels, $[l_1,\ldots, l_n]$, and each one of five "template" vectors corresponding to the five stimulus spectra illustrated in Figure 1. The latter template vectors can be normalized so that the sum of their components equals zero. These vectors are then strictly identical to the optimal weighting vectors described above, where one element is equal to α, and the other four elements to −α/4. An observer who uses weighting patterns different from those will necessarily perform sub-optimally in the considered m-alternative identification task. For instance, a listener who assigns a weight of 1 to the lowest

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

spectral component, and weights of zero to the other components (i.e., the observer only attends to the lowest spectral component) will obviously achieve near-chance performance. More subtle examples of sub-optimal weighting patterns include ones in which the observer gives an unduly large weight to one of the components, but non-zero weights to other components. Clearly, there exists an infinitely large number of sub-optimal decision strategies, corresponding to different vectors of weights. A listener may be using *any* of these weighting patterns. The question is whether the perturbation-correlation approach can be used to uncover which weighting pattern a listener is using.

Extending the Correlational Technique to m-Alternative Identification

In trying to apply the correlational approach in the context of an *m*-alternative identification experiment such as the one described above, one quickly meets a practical problem, which is due to the fact that there are more than two response alternatives. Even though one can still compute a correlation coefficient between the series of random perturbations applied to the levels of the stimulus components across trials, and the numeric ranks (1 to 5) assigned somewhat arbitrarily to the corresponding responses ("S_1" to "S_5"), this calculation is completely irrelevant. Even if the observer behaves optimally and the proportion of responses "S_k" tends to increase with the level of the kth stimulus component this relationship will not be reflected in the correlation coefficient between the series of perturbations and the corresponding series of response ranks across trials. Thus, when the number of possible responses exceeds two, the simple recipe of computing correlation coefficients between stimulus perturbations and numerically coded responses across trials, which forms the basis of the correlational analysis for two-alternative designs (Richards and Zhu, 1994; Lutfi, 1995), no longer works. A second problem relates to the above-mentioned fact that, whereas in the two-alternative case, there is a single weighting pattern, when there are *m* alternatives, the observer or listener may be using up to *m* different weighting patterns; each of these weighting patterns must be estimated.

The correlational analysis of *m*-alternative data, which we propose here, overcomes both of these problems. The analysis includes two steps. First, the trials are sorted into groups, depending on which of the stimuli was presented. Trials on which the same stimulus alternative was presented are grouped together. In our spectral-shape identification example, there were five possible stimulus alternatives. Thus, there are five groups of trials. Each group is further divided into subgroups, corresponding to different signal strengths. In our example, signal strength was represented by the magnitude of the level increment, DL, applied to the target component; the size of this increment determines the strength of the signal that the observer must detect in order to correctly classify the stimulus. Second, for each subgroup, the participant's responses, $R = [r^1, \ldots, r^t]$, are scored for correctness, with a score of 1 assigned to correct responses, and a score of 0 assigned to incorrect responses. The resulting series of scores across trials is denoted as $S = [s^1, \ldots, s^t]$. Finally, for each subgroup, a point-biserial correlation coefficient is computed between the binary scores and the perturbations applied to the level of each frequency component. Note that the correlation is, not directly with the stimulus-component levels, but with the *perturbations* applied to these levels. These level perturbations are analogous to the noise fields used to compute classification-image in visual psychophysics experiments; they do not include the "signal" (i.e., the increment applied to the level of the target component).

Readers familiar with the correlation analysis of two-alternative data will have noticed two important differences between that analysis and the one proposed here. First, whereas in the two-alternative analysis all trials are typically lumped together (Richards and Zhu, 1994; Lutfi, 1995), in the *m*-alternative analysis proposed here, the trials are sorted into blocks based on what stimulus was presented, and the correlations coefficients are computed

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

separately for each trial block. The stimulus-based separation of trials is introduced in the *m*-alternative case because, as mentioned above, the optimal decision strategy involves multiple weighting patterns, which may (and usually do) differ between the different stimulus alternatives. To uncover these weighting patterns, it is useful to separate trials based on which stimulus was presented, before calculating the stimulus-response correlation coefficients. We come back to this important aspect of the analysis in a subsequent section ("Influence of Across-Trial Averaging on Weighting Patterns").

A second important difference between the two analyses relates to the fact that, whereas in the two-alternative case the correlations are between the stimulus variables and the participant's responses (Richards and Zhu, 1994; Lutfi, 1995), here, the correlations are between the stimulus variables and the response *scores*. The reason for this is that with multiple alternatives the responses are no longer binary, and the coding of responses into numbers is, in general, arbitrary. Except perhaps in very special circumstances where the different alternatives can be ordered sensibly along some scale, the ranks assigned to the different responses have no special significance. Thus, correlations involving the response ranks are meaningless. However, one appropriate response-coding strategy is one in which the responses are categorized according to whether or not they match the presented stimulus, so that responses that match the stimulus are coded as 1 while responses that do not match it are coded as 0. This is equivalent to scoring responses for correctness. We come back to this important aspect of the analysis in a subsequent section ("Influence of using Responses instead of Response Scores"), where we also consider alternative approaches based on responses instead of response scores.

In order to illustrate the approach, and confirm that it can yield valid estimates of underlying decision weights, we performed a series of computer simulations using Matlab (The MathWorks, Cambridge, MA). We simulated the behavior of an optimal observer in the *m*-alternative spectral-shape identification task described above. On each simulated trial, one of the five stimuli (S_1, …, S_5) was selected at random (with equal probability). Statistically independent random perturbations were added to the levels of the five stimulus components. These additive perturbations were drawn from a Gaussian distribution with a mean of zero and a standard deviation of 1 dB. The responses of the simulated (or "virtual") observer were determined according to the optimal decision rule described above. Finally, point-biserial correlation coefficients were computed between the series of response scores and the series of level pertubations for each stimulus components. The increment (DL) applied to the level of the target component was varied from 0 to 6 dB in 0.5-dB steps. For each increment level, 50,000 trials were simulated.

Figure 2 displays the point-biserial correlation coefficients between the simulated series of perturbations and the corresponding response scores. The five plots correspond to the five stimuli S_1 to S_5, shown in Figure 1. The results shown in a given plot were obtained using solely the data from the subset of trials on which the indicated stimulus (S_1, S_2, S_3, S_4, or S_5) was presented. In each plot, the decision weights estimated based on the correlation coefficients, are shown as a function of both the frequency-component number (from 1 to 5), and the signal strength (DL). The resulting 3D figures illustrate how the measured weighting functions corresponding to the stimuli, S_1 to S_5, varied as a function of signal strength.

The first observation is that the estimated decision weights were generally positive for the target component (i.e., the first component for signal S_1, the second component for signal S_2, and so on), and negative for the other components. At a given DL, the magnitude of the negative weights was a quarter of the magnitude of the positive weight. This weighting pattern is entirely consistent with the weights corresponding to the optimal decision strategy used in these simulations. As indicated earlier (and detailed in the Appendix), for the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

considered *m*-alternative spectral-shape identification task, the optimal strategy uses as a decision variable the difference between the level of each component and the mean level of the remaining four components in the stimulus. This is equivalent to assigning a positive weight of a to the signal component and a negative weight of $-a/(n-1)$ to each of the remaining $n-1$ components. This outcome demonstrates that the correlational analysis described above can be used to correctly estimate the sign and relative magnitude of the decision weights in an *m*-alternative identification task.

A second important observation on Figure 2 is that the absolute magnitude of the correlation coefficients varied non-monotonically with signal strength, first increasing slightly, and then decreasing markedly toward zero, as DL increased from 0 to 6 dB. This can be understood from Equation 1, which shows that the point-biserial correlation coefficient is proportional to the square-root of the product of the proportions of correct and incorrect responses, $p1$ and $p0$. This product is largest when the two proportions are equal, i.e., when $p1 = p0 = 50\%$. As the proportion of correct (or incorrect) responses moves away from 50% in either direction, the maximum value that can be attained by the point-biserial correlation coefficient decreases.

Importantly, while the measured correlation coefficients varied non-monotonically as a function of signal strength (DL), both their sign and their *relative* magnitude (i.e., the ratio between two correlation coefficients), which define the shape of the weighting patterns, did not vary with signal strength. This indicates that, in principle, decision-weighting patterns in *m*-alternative identification experiments can be estimated regardless of what signal strength was used. However, in practice, one would want to use the signal strength for which the absolute magnitude of the measured correlation coefficients is largest, because the variance of correlation coefficients generally decreases as their magnitude increases. Smaller variance implies greater accuracy in the estimation of the weights, meaning that for the same numbers of trials, an experimenter will be better able to determine whether weights are significantly different from zero and from each other. As mentioned above, the point-biserial correlation coefficient is maximal when the product of correct- and incorrect-response proportions, p_1p_0, is largest, i.e., when the two proportions are equal to 0.5. Accordingly, experimenters should include signal strengths at which participants achieve approximately 50% correct.

Influence of Across-Trial Averaging on Weighting Patterns

An important feature of the optimal decision rule in the *m*-alternative identification task is that it involves different decision weights for evaluating the different stimulus alternatives. This is apparent in Figure 2, which shows that the weighting patterns differ across the various stimulus alternatives. Accordingly, an important step in the correlational analysis of *m*-alternative data involves separating trials based on which stimulus alternative was presented. This should be done before the correlation coefficients between stimulus perturbations and response scores are computed. If the correlation coefficients are calculated with *all* trials pooled together, the outcome will be an averaged version of the weighting patterns corresponding to the different stimulus alternatives.

In order to illustrate this important point, we re-analyzed the data from the Monte Carlo simulations used to produce Figure 2, but this time, pooling all trials together. The resulting "overall weighting function" is shown in Figure 3A. As can be seen, it is almost completely flat. An experimenter looking at this result might be led to the conclusion that the listener assigned equal weights to the five stimulus components. In fact, the overall weighting function obtained by pooling together all trials is essentially an average of the weighting functions for the different stimulus alternatives, which were shown in Figure 2. Because of the symmetry in these weighting functions, their average is flat.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In other cases, the "smeared" weighting patterns obtained by averaging across all trials may not be flat. For instance, suppose the identification experiment described earlier is once again performed, but this time using only three of the five stimulus alternatives – it does not matter which three. The optimal weighting strategy remains the same as described earlier: each stimulus alternative is evaluated by assigning a weight of α (where α is an arbitrary scalar) to the incremented component, and weights of −α/4 to each of the other components. However, because only three stimulus alternatives are now used, and the weighting patterns corresponding to these alternatives do not form a symmetric set, the average weighting pattern obtained when pooling all trials together is no longer flat. This is demonstrated in Figure 3B, which shows the weighting pattern obtained after pooling data across trials on which stimulus S_1, S_2, or S_3 was presented, and excluding other trials. Note that, unlike in previous plots, the weighting pattern is only shown here for a single signal strength (DL = 0 dB), rather than as a function of singal strength; hence the 2D format of this plot. As in previous plots, although the magnitude of the measured correlation coefficients varied as a function of DL, the relative magnitudes and signs of the coefficients remained constant. As can be seen, the resulting pattern exhibits positive weights on the first three components, and negative weights on the last two. This pattern might suggest to an unwary experimenter that the underlying decision strategy consisted in subtracting the mean level of the last two components from the mean level of the first three. This strategy is completely useless for the task at hand, and does not correspond to that used by our virtual observer. The pattern of correlation coefficients shown in Figure 3B is just the mean of the optimal weighting patterns associated with the three stimulus alternatives that were arbitrarily selected in the simulated experiment.

These observations illustrate some of the pitfalls that can result from pooling data indiscriminately across trials when analyzing the results of an *m*-alternative identification experiment using the correlational technique.

Influence of using Responses instead of Response Scores

The weighting patterns in Figure 2 are based on the point-biserial correlation coefficients between the perturbations applied to the stimulus levels and the binary response scores. As we pointed out earlier, this differs from the correlational analysis for two-alternative data, where the correlation is with the *responses* rather than with the response *scores*. As we also explained earlier, when the number of response alternatives is larger than two, computing the correlation between the perturbed stimulus variables and the (*m*-ary) coded responses is not a sensible thing to do. However, scoring for correctness is not the only way to transform the responses into binary values, whose correlation with stimulus variables may yield estimates of the underlying decision weights. Another approach is to transform the responses in such a way that all responses corresponding to a given alternative, say, S_3, are coded as 1, while all the other responses (i.e., "S_1", "S_2", "S_4", and "S_5") are coded as 0, regardless of which stimulus alternative was actually presented. This approach may be regarded as closer to that used in the correlational or classification-image analyses for the two-alternative case, in which the *responses* (and not the actually presented stimulus) are regarded as the important variable.

The results of this alternative analysis are illustrated in Figure 4A. This plot uses the same 3D format as that of Figures 2 and 3A: the estimated weights for the five spectral components are shown as a function of the signal strength (DL). As in Figure 3A, and unlike for Figure 2, here, only the weighting patterns for alternative S_3 are shown. The key difference between this plot and that labeled "S_3" in Figure 2 is that, here, "S_3" refers to the *response* rather than to the *stimulus.* The weights shown in this plot were estimated based on the correlation coefficients between the stimulus perturbations and a binary vector with

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

components equal to 1 whenever the response was "S_3", and to 0 otherwise, across *all* trials. The resulting weighting patterns appear to not differ dramatically from those shown in the "S_3" panel of Figure 2. In particular, they exhibit a peak for the third stimulus component, and at a given DL, the magnitude of the weight on the third component is approximately four times larger than that of the negative weights corresponding to the other components. This is consistent with the decision rule used in these simulations. However, comparing Figure 2 and Figure 4A, it can be seen that the weights approach zero more rapidly as a function of signal strength in the latter than in the former. An intuitive explanation for this effect is that not taking into account which stimulus was actually presented results in more unexplained variability in the measured relationship between stimuli and decisions; a larger amount of unexplained variability implies a lower correlation coefficient. This shows that, although it is possible to estimate decision weights in an *m*-ary identification task based on the responses rather than on the response scores, this approach is less statistically efficient than that based on the response scores.

As we pointed out in the Introduction, for two-alternative data, the correlational analysis is closely related to the classification-image analysis. The relationship is made explicit by Equation 1. As mentioned earlier, this alternative formula for the point bi-serial correlation coefficient between the perturbed stimulus variables and the observer's binary responses, which forms the basis of the correlational analysis, is directly related to the difference between the expected values of the random stimulus variable conditioned on the response (0 or 1), which is at the heart of the classification-image analysis. This begs the question of whether a similar correspondence exists between the outcome of the correlational and classification-image analyses in the *m*-alternative identification case, where $m>2$. To address this question, we re-analyzed the Monte Carlo simulation data obtained earlier, this time using a modified version of the classification-image analysis, which was adapted specifically to deal with more than two alternatives. First, the perturbations applied to the levels of the stimulus components were (separately) averaged across all trials on which the response was "S_3", regardless of which stimulus was actually presented. Second, the perturbations were averaged across all trials on which the response was different from "S_3", again, regardless of which stimulus was actually presented. To facilitate comparison with Figure 2C, the results of these analyses were scaled in such a way that, for the null signal strength (DL = 0 dB), the estimated "weight" of the middle component was the same in the two figures.

The results are shown in Figure 4, panels B and C, respectively. The first thing to note is that, like the weighting patterns shown in Figures 2C and 4A, the results of the analysis using only trials on which the response was "S_3" (Figure 4B) show a peak on the middle component. In addition, the weights estimated using this analysis have the same relative magnitudes as those obtained using the above-described analyses. The pattern of weights estimated based on trials on which the response was different from "S_3" (Figure 4C) is just the negative of that in Figure 4B. One can easily recover the correct weighting pattern by inverting the signs of these weight estimates.

These results reveal that an analysis based on differences between response-conditional means of the stimulus perturbations can yield valid decision-weight estimates in an *m*-alternative identification task of the type considered here. This outcome is worth emphasizing because, to our knowledge, this provides the first demonstration that a variant of the classification-image analysis can be applied in the context of a task involving more than two alternatives. In this variant, the decision weights are estimated by computing the difference between the average stimulus across all trials on which a given response was produced, and the average stimulus across all remaining trials.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

As a final remark, it is worth pointing out that as the signal strength approaches zero, the distinction between the score- and response-based estimation procedures that we considered in this section vanishes. Obviously, when there is no signal (which is the case when DL = 0 dB in our example spectral-shape identification experiment), the different stimulus alternatives are all identical, and classifying trials based on which stimulus was presented, or on response correctness, becomes completely arbitrary. Under such circumstances, all the analyses described above are equivalent.

Influence of Internal Noise

In the above-described simulations, the stimulus perturbations were the only source of random variability. The sensory observations were assumed to be accurate reflections of the levels of the stimulus components. In reality, sensory observations are almost invariably corrupted to some degree by internal noise. Internal noise subsumes random variability in neural responses as well as other sources of observer-response variability unrelated to the stimulus, such as attentional lapses. Since the correlation between stimulus perturbations and response scores does not capture variability in the observer's responses that is unrelated to the stimulus perturbations, internal noise is expected to reduce the maximum attainable magnitude of the measured correlation coefficients.

To examine the influence of internal noise on correlation-based estimates of decision weights, we performed a series of Monte Carlo simulations similar to those described above, except that internal noise was included. It was simulated by adding random perturbations to the level of each stimulus component. These "internal" perturbations were added on top of the "external" perturbations that were applied to the stimulus on each presentation. Following a long-standing tradition in psychophysical signal detection theory (Green & Swets, 1966), the internal noise was assumed to be zero-mean Gaussian. Based on data in the psychoacoustics literature, which indicate just noticeable differences (JNDs) of about 1 dB in level-discrimination tasks (e.g., Zwicker, 1970), the standard deviation of the Gaussian internal-noise distribution was initially set to one, and was then equal to the standard deviation of the external perturbations. However, in subsequent simulations, the standard deviation of the internal noise was systematically varied in order to study the influence of this model parameter on the estimated weighting patterns. For the simulations described in the current section, the internal noise was uncorrelated across the five stimuli[2]; the influence of internal-noise correlations is considered in a subsequent section.

The correlation coefficients measured based on the results of simulations including internal noise with a 1-dB standard deviation are shown in Figure 5A. As expected, the magnitude of the correlation coefficients was reduced, compared to the case where internal noise was absent (Figure 2C). More surprisingly, at relatively large signal strengths (i.e., large DLs), the magnitude of the correlation coefficients was actually larger, compared to the case with no internal noise. This outcome is related to the fact that the correlation coefficients are proportional, not only to the underlying decision weights, but also to the square-root of the product of the proportions of correct and incorrect responses (see Equation 1). As mentioned earlier, this product decreases monotonically as the proportion of correct responses increases above 50%. As a result, increases in signal strength lead to a *decrease* in the measured correlation coefficients toward zero. Internal noise partly counteracts this effect by reducing

[2]Under the optimal decision rule, any internal noise that is common to all frequency components (or channels) will not affect the correlation between the responses and the stimulus perturbations that are independent across the components, because any common noise terms are canceled in the decision variable. Under a non-optimal decision rule, however, common noise terms can exert their own influence on the responses, effectively reducing the potential influence of the perturbations on the responses. Therefore, the presence of both independent and common terms of internal noise can reduce the maximum value that can be attained by the point-biserial correlation coefficient between the perturbations and the responses.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the proportion of correct responses back toward 50%. However, by and large, the pattern of measured correlation coefficients across the five stimulus components still resembles the predicted pattern of weights under the assumed decision rule. This indicates that, even in the presence of internal noise having a magnitude comparable to the size of the perturbations, the correlation analysis can be used to gain insight into individual decision strategies in an *m*-alternative identification task.

Situations in which internal noise can severely compromise the experimenter's ability to estimate decision weights are those in which the magnitude of the internal noise is substantially larger than that of the stimulus perturbations. Figure 5B shows the weighting patterns obtained in simulations involving different amounts of internal noise. The standard deviation of the internal noise was set to 0, 1, 2, 4, 8, or 16 times that of the stimulus perturbations, which remained fixed at 1 dB. The signal strength (DL) was also fixed at 0 dB in these simulations; this corresponds to the case of catch trials, where the five stimulus components had the same level, on average. Using catch trials here simplifies the interpretation of the results by avoiding an interaction between signal strength (DL) and correlation coefficients. The shaded area represents the 95% confidence interval around a correlation coefficient of 0 computed based on 2,000 trials; this relatively small number of trials was used in estimating the confidence interval here because it is more typical of the number of trials likely to be performed in actual psychophysical experiments than that used in estimating the weights – which was chosen very large to ensure asymptotic convergence. Measured correlation coefficients that fall inside the area are deemed not significantly different from zero. It can be seen that, when the standard deviation of the internal noise is more than four times larger than that of the perturbations, the correlation coefficients corresponding to the non-signal components are not significantly different from zero. Upon seeing this pattern of results, an experimenter might conclude erroneously that the observer assigned zero weights to these components. However, unless very small perturbations are used, internal noise with a magnitude more than four times larger than that of the external perturbations would be highly unusual in most psychophysical experiments. Other than using reasonably large perturbations, a simple way in which experimenters can guard against this problem is to include signal strengths that will yield proportions of correct responses close to 50%.

Influence of Correlations between the Observations

Internal noise is not the only factor that can hamper accurate estimation of decision weights using the correlational technique. Another factor relates to the existence of correlations between the sensory observations. Such correlations can result from nearby stimulus components exciting partially overlapping portions of the sensory epithelium or neural populations, so that they are not processed independently. In our example, this would occur if the frequency separation between nearby spectral components in the stimuli shown in Figure 1 were smaller than the bandwidth of peripheral auditory filters, so that auditory channels that respond to one component also respond (albeit to a lesser extent) to the other. As a result, the variations in the output of the corresponding auditory channels due to the stimulus perturbations will be partly correlated, even if the physical perturbations applied to the individual components are uncorrelated. In general, statistical dependencies among the sensory observations will influence the measured correlation coefficients between stimulus perturbations and response scores. Depending on their structure, these dependencies between observations can reduce the magnitude of the measured correlation coefficients uniformly, or alter the relative magnitude of these coefficients (by reducing some more than others), thereby distorting the observed weighting pattern.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

To illustrate this effect, we performed simulations in which the sensory observations evoked by each stimulus component were computed as a weighted sum of the level of that component and the levels of the other four components. The four non-target components were assigned the same weight, which was set to 0, 0.2, 0.4, 0.6, 0.8, or 1 times the weight of the target component; these values correspond to correlation coefficients of 0, 0.45, 0.78, 0.93, 0.99, and 1, between the target and non-target components, respectively. This correlation structure, which corresponds to equal correlation between the sensory observations evoked by target and non-target components regardless of their distance along the frequency axis, was chosen here for illustration purposes.[3] This correlation structure preserves the symmetry of the weighting patterns with respect to the different stimulus alternatives: the expected effect of the correlation is to reduce the magnitude of the measured correlation coefficients uniformly, while leaving their sign unchanged, yielding estimated weighting patterns that are qualitatively similar to (and can be directly compared with) those shown in previous figures. In these simulations, the signal strength (DL) was again set to 0 dB, the magnitude of the stimulus perturbations to 1 dB, and the standard deviation of the internal noise also to 1 dB. The measured weighting patterns are shown in Figure 6. As for Figure 5b, the shaded area around zero correlation represents the 95% confidence interval (based on 2,000 trials). As expected, as the correlation between the sensory observations increases, the magnitude of the measured correlation coefficients decreases. When the correlation between the observations evoked by the target and non-target components exceeds 0.93, the estimated weights for the non-signal components can no longer be differentiated reliably from zero. A correlation of 0.93 or more between sensory observations should not be expected in experiments in which the stimulus components to which the perturbations are applied are relatively well separated in space (for vision) or frequency (for hearing). The conclusion is that, except perhaps in situations where very high correlations between a subset of the sensory observations are expected, the experimenter's ability to estimate decision weights in an *m*-alternative task using the correlational approach described here should not be seriously compromised.

Influence of Departures from Normality of the Perturbations

So far, we have assumed that the stimulus perturbations had a Gaussian distribution. In some situations, the experimenter may have no choice other than using perturbations that are not strictly Gaussian. This is likely to occur in experiments in which the stimulus variable to which the perturbations are applied is bounded, or can only take on discrete values. Thus, an important question for experimenters is whether the correlational technique described here can still be used to correctly estimate decision weights in situations where the perturbations are not Gaussian.

In order to address this question, we performed Monte Carlo simulations similar to those performed earlier to produce the results shown in Figure 5, except that this time the random level perturbations applied to the individual stimulus components were drawn from a variety of distributions rather than just the Gaussian distribution. The distributions were generated using the following equation:

$$p(x)=\kappa\, e^{-\left|\frac{x}{\sqrt{2}}\right|^{\beta}} \tag{4}$$

[3]Frequency selectivity curves in the auditory system are usually v-shaped, so that nearby frequency components are more likely to evoke correlated neural responses than components spaced farther apart. The correlation structure chosen here corresponds to a situation in which the non-target components are relatively far away from the target, and the auditory filter centered on the target frequency attenuates the level of all the non-target components approximately uniformly.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

where $|x|$ denotes the absolute-value of x, β is a positive exponent (β>0), which controls the kurtosis of the distribution, and κ is a normalization factor, which depends on β and ensures that the distribution sums to 1. Five β values were used in the simulations: 1, $\sqrt{2}$, 2, $2\sqrt{2}$, 4. The corresponding probability distributions are shown in Figure 7. For $\beta = 1$, $p(x)$ is the Laplace distribution; for $\beta = 2$, it is the Gaussian. As β keeps increasing, $p(x)$ tends toward the uniform. Although the decision strategy involving a positive weight of α on the target component and negative weights of −α/4 on each non-target component is no longer optimal when the observations are not Gaussian, it was kept in the current simulations to facilitate comparisons with the results obtained previously. Ideally, the ability of the correlational technique to uncover the underlying decision strategy should be independent of the specific form of the decision rule.

The estimated weighting patterns for the different distributions defined above are shown in Figure 8. The weight estimates were computed using trials on which the level increment (DL) was applied to the middle component ($i = 3$). For the upper panel, DL was set to 0 dB; for the lower panel, it was set to 1.5 dB. As can be seen, the weighting patterns obtained in the absence of signal (Figure 8A) are essentially identical, all reflecting the optimal decision rule, regardless of the distributions of the perturbations. With a 1.5-dB signal (Figure 8B), the weighting patterns obtained using different distributions of perturbations still resemble each other, but they do not overlap as closely; depending on which distribution of perturbations was used, the estimated weights are shifted and stretched vertically to different extents. Qualitatively, these weighting patterns displaying a positive weight on the signal component, and smaller negative weights on the non-signal components, are consistent with the underlying decision strategy. However, the relative magnitudes of the estimated weights for the signal and non-signal components decreased systematically as the kurtosis of the distribution of perturbations increases (or decreasing β). The only distribution for which the inferred weighting pattern accurately reflected the underlying (true) decision weights was the Gaussian.

The difference between results obtained using Gaussian and non-Gaussian distributions of perturbations is illustrated in Figure 9 in a different way. This plot shows the sum of the measured weights across the five frequency components as a function of signal strength (DL) for the different distributions of perturbations. In addition to the distributions shown in Figure 7, a uniform distribution was also included in these simulations to illustrate what happens in the limit where the distribution of perturbations has the smallest possible kurtosis. [4] The optimal decision rule requires that the sum of the weights be zero, independent of signal strength. Looking at Figure 9, it can be seen that the Gaussian (dark dashed line) is the only distribution for which this was true.[5] For all other distributions of perturbations, the sum of weights varied non-monotonically with signal strength. Departures from zero were largest at medium signal strengths (i.e., for DLs comprised between 1 and 4 dB), and were positive for platykurtic or "sub-Gaussian" distributions, i.e., distributions with a lower kurtosis than the Gaussian, such as the uniform distribution (solid line), but negative for leptokurtic or "super-Gaussian" distributions (i.e., distributions with a higher kurtosis than the Gaussian), such as the Laplace ($\beta = 1$, grey dashed line). There is a simple explanation why the sum of the weights approaches zero at large signal strengths for all of these distributions: when the signal is easily detectable, the observer's responses are predominantly correct ($p1$ close to 1); as a result, the correlations between the perturbations

[4]The range of the uniform distribution used in these simulations was set to 6 dB, which corresponds to an area of 99.8% for the Gaussian distribution of perturbations with a standard deviation of 1 dB.

[5]The small fluctuations in those curves are due to the use of a finite number of trials in the simulations. In the simulations reported in this section, the number of trials was set to 500,000, so that the curves would be relatively smooth. In an actual psychophysical experiment, the number of trials would likely be substantially smaller (typically, less than 5,000). Therefore, the variability in the measured weights and their sum across stimulus components would be larger.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

and the scores are close to zero, and so is their sum. The reason why the sum of the weights obtained with non-Gaussian perturbations approaches zero at very small signal strengths remains unclear.

These observations reveal that results obtained using the correlational technique in *m*-alternative experiments involving non-Gaussian rather than Gaussian perturbations should be interpreted with caution. The failure of the correlation coefficients to correctly reflect the underlying decision rule when the perturbations are not Gaussian stems from the fact that, unless the observations have a Gaussian distribution, the correlation coefficients between perturbations and responses are no longer related linearly to the underlying decision weights. When there is no control over the distributions for the stimulus variation, other things being equal, trials on which the signal strength was lowest are least likely to be affected by departures from Gaussianity. The weighting patterns estimated from these trials will most likely reflect the underlying decision rule.

Influence of Non-Linearities in the Stimulus-to-Decision Mapping

An important issue in applications of the psychophysical reverse-correlation approach relates to the presence of non-linearities in the mapping from the stimulus space to the decision space. Non-linearities may arise at the level of the physical-to-sensory transformation, or at the level of the transformation between the observations and the decision variable. The former typically reflect non-linear processes in the sensory periphery, such as cochlear compression. The effect of such "peripheral" non-linearities on the estimation of decision weights has been considered in earlier publications (Abbey & Eckstein, 2002; Dai, 2000; Dai, Nguyen, Green, 1996; Murray et al., 2002; Neri, 2004; Richards, 2002; Richards, Tang, & Kidd, 2002; Tjan & Nandy, 2006). The second type of non-linearity arises at the level of the relationship between the sensory observations and the decisions. This type of non-linearity is inherent to the decision model for the *m*-alternative case, because as we showed earlier the optimal decision rule for *m*-alternative identification involves finding the maximum of *m* likelihood functions (cf. "Decision Model and Optimal Decision Rule for m-Alternative Identification" and Appendix A). The maxima of the different likelihood functions may occur at different points in the observation space. If so, the selection of the largest of these maxima (i.e., the absolute maximum) over the entire decision space may not, in general, be modeled by a single linear function. The five-alternative spectral-shape identification example, which we used in the simulations, represents a special case because, in this example, the decision surface actually consists of *m* planes with different orientations. In practice, experimenters may encounter situations in which the decision surface does not consist of planes, but rather of curved surfaces. In order to understand how one may be able to deal with such situations, let us consider how non-linearities in the relationship between the stimulus variable(s) and the decision variable can affect the estimation of decision weights using the correlational approach. In the presence of non-linearities, the decision variable, *Z*, is no longer simply a weighted sum of the observations, and Equation 2 must be replaced by:

$$Z = f(y_1, y_2 \ldots, y_n), \tag{5}$$

with *f* a non-linear function. In this case, the surface traced by the decision statistic, *Z*, in the decision space is a curved surface rather than a plane. The intersections of this decision surface with the two-dimensional planes formed by one of the observation axes and the decision-statistic axis are curves rather than lines. Therefore, unlike for the linear decision model, the decision surface for the non-linear decision model cannot in general be characterized completely by a single set of *n* coefficients corresponding to the decision

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

weights. However, the correlational approach can still be used to probe individual decision strategies, because stimulus-response correlations provide estimates of the first-order partial derivatives of the decision statistic with respect to each of the observation axes at a specific point in the observation space. This follows directly from the realization that, if the function $f(y_i)$, which relates the random variable y_i (denoting observations along the ith dimension of the observation space) to the decision statistic Z, is expanded using a Taylor series at some point, $y_i = a$, as follows:

$$f(y_i)=f(a)+(y_i-a)\frac{df}{dy_i}\Big|_{y_i=a}+(y_i-a)^2\frac{d^2f}{2dy_i^2}\Big|_{y_i=a}+(y_i-a)^3\frac{d^2f}{6dy_i^3}\Big|_{y_i=a}+\ldots \tag{6}$$

the first non-constant term in the resulting series is a linear function of the observations. Provided that the function $f(y_i)$ is reasonably smooth in the vicinity of a, the higher-order terms in the right-hand side of Equation 6 can be ignored, and $f(y_i)$ may be approximated as a linear function, as follows:

$$f(y_i) \cong f(a)+(y_i-a)\frac{df}{dy_i}\Big|_{y_i=a} \tag{7}$$

In this context, the first-order partial derivatives provide estimates of the decision weights at a specific point in the observation space. To distinguish these "local" weights from the constant weights of the linear model, $w_1, \ldots, w_n$, and emphasize that their value depends upon the considered point in the observation space, we denote them as $\omega_1(Y), \ldots, \omega_n(Y)$.

Empirically, the local weights, $\omega_1(Y), \ldots, \omega_n(Y)$, can be estimated by using a relatively small range of perturbations around a fixed stimulus level, chosen to correspond to a point of interest in the decision space. In theory, the whole decision surface can be explored using this approach, by sampling a sufficiently large number of locations in the observation space (i.e., by using multiple stimulus levels, around which a narrow range of perturbations is applied). In practice, the large number of trials required to explore multiple points in the stimulus space prohibits exhaustive exploration. Fortunately, in some experiments, the constraints imposed by the design are such that the decision surface must exhibit some regularity. For instance, the surface may be highly symmetric with respect to one or several axes in the decision space. This was the case for our spectral-shape identification example: because of the way in which the spectra were constructed, the five stimulus alternatives form a "symmetric" or "balanced" set. The experimenter can take advantage of such symmetry to determine some key features of an observer's decision strategy without having to sample an impractically large number of points in the observation space.

Influence of number of trials

The large number of trials required to obtain interpretable results is a general concern in practical applications of the correlational and classification-image techniques. The problem of impractically large numbers of trials becomes particularly acute as the number of stimulus-response alternatives increases, because each additional alternative requires an increase in the number of trials in order to achieve the same accuracy as with just two alternatives. It is not atypical for two-alternative classification-image experiments in vision to involve 10,000 trials or more per observer. To achieve the same estimation accuracy in comparable m-alternative experiments, one would need m times 5,000 trials. However, Watson (1998) obtained meaningful classification images in a three-alternative letter identification task using only 5,888 trials in one observer. Moreover, successful applications

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of the correlational approach in two-alternative auditory detection or discrimination tasks have been achieved using even smaller numbers of trials (e.g., 1,000 trials for each estimated weight in Dai, 2000; 25 to 1000 trials for each estimated weight in Dai et al., 1996). This may be because of the reduced dimensionality of auditory stimuli compared to visual stimuli; whereas an image (and the corresponding classification image) typically contains thousands of pixels, the sounds used in auditory correlational experiments (and the corresponding weighting patterns) usually involved less than ten frequency components, or frequency bands.

In order to investigate how the decision-weight estimates in our example spectral-shape identification task might depend on the number of trials, we computed Pearson correlation coefficients between the theoretical (optimal) weighting patterns, and the weighting patterns estimated based on the results of Monte Carlo simulations similar to those described above, but involving fewer trials.[6] For these simulations, a Gaussian distribution of perturbations was used, and internal noise with a standard deviation of 1 dB was included. The results are illustrated in Figure 10. This shows the mean correlation coefficient between the theoretical and measured weighting patterns (across 1,000 simulations) as a function of signal strength (DL), with the number of trials as the parameter. Note that the y-axis scale ranges from 0.8 tom 1. Even when only 50 trials were used, the average correlation between the theoretical and measured weighting pattern was already above 0.85 on average. With 800 trials, the correlation was already close to 1 at small signal strengths (around 0 dB), and only started to drop rapidly for signal strengths larger than about 4 dB – at which the proportion of correct responses of the simulated observer was well above chance. These observations suggest that for a relatively simple, five-alternative, five-stimulus-component identification task such as the one used as example in this note, it may be possible to obtain interpretable weighting patterns using a relatively small number of trials.

Conclusions

Uncovering the internal stimulus representations and decision strategies that determine observers' or listeners' performance in various visual or auditory perception tasks is a fundamental goal of psychophysical research. Psychophysical reverse-correlation methods such as the classification-image and correlational techniques are ideally suited for achieving this goal. Although these methods have been the object of rapidly increasing theoretical and applied interest in recent years, thus far, the question of their applicability in experiments involving more than two response alternatives had not been considered in any great detail.

In this note, some of the most important theoretical and practical problems posed by applications of the "correlational" and "classification image" methods (the two most popular psychophysical reverse-correlation methods) in identification experiments involving more than two response alternatives were explained and addressed. A modified correlational analysis was described, wherein "weighting patterns" (or "perceptual templates") corresponding to different stimulus-response alternatives were estimated using point-biserial correlation coefficients between stimulus perturbations and response scores within stimulus-specific trial blocks. Computer (Monte Carlo) simulations were performed to test the proposed approach, and compare it to alternative approaches. The results revealed that an approach based on *response scores* was superior to alternative methods (including a modified "image classification" analysis) based on *responses*. Reasons for this superiority include the fact that response scores implicitly contain information about which stimulus

[6]A similar analysis was performed by Murray et al. (2005) to evaluate the efficiency of observers' weighting strategies in a visual two-alternative identification task; they used the cross-correlation between the ideal template and an observer's classification image as an estimate of the efficiency of that observer's decision strategy.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

was presented, in addition to information about which response was chosen. The pitfalls of indiscriminate across-trial averaging and arbitrary selection of a subset of stimuli were clearly demonstrated.

The simulation results further revealed that the proposed correlational analysis yielded valid estimates of relative decision weights even when relatively large internal noise (compared to the magnitude of the external perturbations), or relatively high correlations between the sensory observations, were present. Various non-Gaussian distributions of stimulus perturbations were also tested, and were found to affect, but not prevent, the estimation of the underlying decision weights. Overall, these findings indicate that the correlational approach proposed here is applicable in principle under a fairly wide range of conditions.

Finally, the question of how non-linearities in the stimulus-to-response mapping may impact the estimation of decision weights was considered. A general answer to this question is that, even when there are more than two response alternatives, correlations between stimulus perturbations and response scores can be used to recover the decision strategy reasonably accurately, as long as the decision variable remains approximately linear within the range spanned by the stimulus perturbations.

It is hoped that the present note will facilitate as well as promote future applications of the psychophysical reverse-correlation approach in visual and auditory experiments involving more than two response alternatives.

Acknowledgments

This work was supported by the University of Arizona and the National Institutes of Health, National Institute on Deafness and other Communication Disorders Grant R01 DC 05216.

The authors are grateful to Mark Stellmack, whose insightful questions and constructive comments prompted some of the analyses described in the text, and to Associate Editor, Isabel Gauthier, two anonymous reviewers, and Joshua McDermott for helpful suggestions.

References

Abbey CK, Eckstein MP. Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. Journal of Vision. 2002; 2:66–78. [PubMed: 12678597]

Ahumada AJ Jr. Classification image weights and internal noise level estimation. Journal of Vision. 2002; 2:121–131. [PubMed: 12678600]

Ahumada AJJ. Perceptual classification images from vernier acuity masked by noise. Perception. 1996; 26:18.

Ahumada AJJ, Beard BL. Response classification images in vernier acuity. Investigations in Opthalmology and Visual Science. 1998; 40:S572.

Ahumada AJJ, Lovell J. Stimulus features in signal detection. Journal of the Acoustical Society of America. 1971; 49:1751–1756.

Ahumada AJJ, Marken R, Sandusky A. Time and frequency analyses of auditory signal detection. Journal of the Acoustical Society of America. 1975; 57:385–390. [PubMed: 1117090]

Alexander JM, Lutfi RA. Informational masking in hearing-impaired and normal-hearing listeners: sensation level and decision weights. Journal of the Acoustical Society of America. 2004; 116:2234–2247. [PubMed: 15532655]

Apoux F, Bacon SP. Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise. Journal of the Acoustical Society of America. 2004; 116:1671–1680. [PubMed: 15478433]

Beard BL, Ahumada AJ Jr. Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. Journal of the Optical Society of America A. 1999; 16:755–763.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Berg BG. Analysis of weights in multiple observation tasks. Journal of the Acoustical Society of America. 1989; 86:1743–1746.

Berg BG, Green DM. Spectral weights in profile listening. Journal of the Acoustical Society of America. 1990; 88:758–766. [PubMed: 2212301]

Bouet R, Knoblauch K. Perceptual classification of chromatic modulation. Visual Neuroscience. 2004; 21:283–289. [PubMed: 15518201]

Brenner E, Granzier JJ, Smeets JB. Perceiving colour at a glimpse: the relevance of where one fixates. Vision Research. 2007; 47:2557–2568. [PubMed: 17692885]

Buus S. Temporal integration and multiple looks, revisited: weights as a function of time. Journal of the Acoustical Society of America. 1999; 105:2466–2475. [PubMed: 10212427]

Buus S, Zhang L, Florentine M. Stimulus-driven, time-varying weights for comodulation masking release. Journal of the Acoustical Society of America. 1996; 99:2288–2297. [PubMed: 8730075]

Calandruccio L, Doherty KA. Spectral weighting strategies for sentences measured by a correlational method. Journal of the Acoustical Society of America. 2007; 121:3827–3836. [PubMed: 17552730]

Calandruccio L, Doherty KA. Spectral weighting strategies for hearing-impaired listeners measured using a correlational method. Journal of the Acoustical Society of America. 2008; 123:2367–2378. [PubMed: 18397039]

Chauvin A, Worsley KJ, Schyns PG, Arguin M, Gosselin F. Accurate statistical tests for smooth classification images. Journal of Vision. 2005; 5:659–667. [PubMed: 16356076]

Dai H. On the relative influence of individual harmonics on pitch judgment. Journal of the Acoustical Society of America. 2000; 107:953–959. [PubMed: 10687704]

Dai H, Green DM. Auditory intensity perception: successive versus simultaneous, across-channel discriminations. Journal of the Acoustical Society of America. 1992; 91:2845–2854. [PubMed: 1629477]

Dai H, Nguyen Q, Green DM. Decision rules of listeners in spectral-shape discrimination with or without signal-frequency uncertainty. Journal of the Acoustical Society of America. 1996; 99:2298–2306. [PubMed: 8730076]

Dai H, Versfeld NJ, Green DM. The optimum decision rules in the same-different paradigm. Perception & Psychophysics. 1996; 58:1–9. [PubMed: 8668509]

Doherty KA, Lutfi RA. Spectral weights for overall level discrimination in listeners with sensorineural hearing loss. Journal of the Acoustical Society of America. 1996; 99:1053–1058. [PubMed: 8609288]

Doherty KA, Lutfi RA. Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners. Journal of the Acoustical Society of America. 1999; 105:1831–1840. [PubMed: 10089606]

Doherty KA, Turner CW. Use of a correlational method to estimate a listener's weighting function for speech. Journal of the Acoustical Society of America. 1996; 100:3769–3773. [PubMed: 8969478]

Dotsch R, Wigboldus DH, Langner O, van Knippenberg A. Ethnic out-group faces are biased in the prejudiced mind. Psychological Science. 2008; 19:978–980. [PubMed: 19000205]

Durlach NI, Braida LD, Ito Y. Towards a model for discrimination of broadband signals. Journal of the Acoustical Society of America. 1986; 80:63–72. [PubMed: 3745667]

Eckstein MP, Ahumada AJ Jr. Classification images: a tool to analyze visual strategies. Journal of Vision. 2002; 2:1. [PubMed: 12678593]

Eckstein MP, Shimozaki SS, Abbey CK. The footprints of visual attention in the Posner cueing paradigm revealed by classification images. Journal of Vision. 2002; 2:25–45. [PubMed: 12678595]

Gilbert G, Micheyl C. Influence of competing multi-talker babble on frequency importance functions for speech measured using a correlational approach. Acta Acustica United with Acustica. 2005; 91:145–154.

Gold JM, Murray RF, Bennett PJ, Sekuler AB. Deriving behavioural receptive fields for visually completed contours. Current Biology. 2000; 10:663–666. [PubMed: 10837252]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Gosselin F, Schyns PG. Bubbles: a technique to reveal the use of information in recognition tasks. Vision Research. 2001; 41:2261–2271. [PubMed: 11448718]

Green DM. Consistency of auditory detection judgments. Psychological Review. 1964; 71:392–407. [PubMed: 14208857]

Green, DM. Profile analysis. Auditory intensity discrimination. New York: Oxford University Press; 1988.

Green DM, Birdsall TG. Detection and recognition. Psychological Review. 1978; 85:192–206.

Green, DM.; Swets, JA. Signal Detection Theory and Psychophysics. New York: Krieger; 1966.

Keane BP, Lu H, Kellman PJ. Classification images reveal spatiotemporal contour interpolation. Vision Research. 2007; 47:3460–3475. [PubMed: 18053850]

Knill DC. Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. Vision Research. 1998; 38:2635–2656. [PubMed: 12116709]

Kornbrot, DE. Point biserial correlation. In: Everitt, BS.; Howell, DC., editors. Encyclopedia of Statistics in Behavioral Science. New York: Wiley; 2005.

Kortekaas R, Buus S, Florentine M. Perceptual weights in auditory level discrimination. Journal of the Acoustical Society of America. 2003; 113:3306–3322. [PubMed: 12822803]

Kurki I, Hyvarinen A, Laurinen P. Collinear context (and learning) change the profile of the perceptual filter. Vision Research. 2006; 46:2009–2014. [PubMed: 16473386]

Levi DM, Klein SA. Classification images for detection and position discrimination in the fovea and parafovea. Journal of Vision. 2002; 2:46–65. [PubMed: 12678596]

Levi DM, Klein SA. Noise provides some new signals about the spatial vision of amblyopes. Journal of Neuroscience. 2003; 23:2522–2526. [PubMed: 12684436]

Levi DM, Klein SA, Chen I. What is the signal in noise? Vision Research. 2005; 45:1835–1846. [PubMed: 15797773]

Levi DM, Klein SA, Chen I. The response of the amblyopic visual system to noise. Vision Research. 2007; 47:2531–2542. [PubMed: 17697689]

Levi DM, Klein SA, Chen I. What limits performance in the amblyopic visual system: seeing signals in noise with an amblyopic brain. Journal of Vision. 2008; 8:1–23.

Lutfi RA. Correlation coefficients and correlation ratios as estimates of observer weights in mutliple-observation tasks. Journal of the Acoustical Society of America. 1995; 97:1333–1334.

Lutfi RA. Auditory detection of hollowness. Journal of the Acoustical Society of America. 2001; 110:1010–1019. [PubMed: 11519569]

Lutfi RA, Jesteadt W. Molecular analysis of the effect of relative tone level on multitone pattern discrimination. Journal of the Acoustical Society of America. 2006; 120:3853–3860. [PubMed: 17225412]

Lutfi RA, Oh EL. Auditory discrimination of material changes in a struck-clamped bar. Journal of the Acoustical Society of America. 1997; 102:3647–3656. [PubMed: 9407656]

Lutfi RA, Wang W. Correlational analysis of acoustic cues for the discrimination of auditory motion. Journal of the Acoustical Society of America. 1999; 106:919–928. [PubMed: 10462797]

Macmillan, NA.; Creelman, CD. Detection theory: A user's guide. 2. Mahwah, NJ: Erlbaum; 2005.

Mangini MC, Biederman I. Making the ineffable explicit: estimating the information employed for face classification. Cognitive Sciences. 2004; 28:209–226.

Mareschal I, Dakin SC, Bex PJ. Dynamic properties of orientation discrimination assessed by using classification images. Proceedings of the National Academy of Sciences USA. 2006; 103(13): 5131–5136.

Mareschal I, Morgan MJ, Solomon JA. Contextual effects on decision templates for parafoveal orientation identification. Vision Research. 2008 in press.

Martin-Malivel J, Mangini MC, Fagot J, Biederman I. Do humans and baboons use the same information when categorizing human and baboon faces? Psychological Science. 2006; 17:599–607. [PubMed: 16866746]

Mehr MA, Turner CW, Parkinson A. Channel weights for speech recognition in cochlear implant users. Journal of the Acoustical Society of America. 2001; 109:359–366. [PubMed: 11206164]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Michel MM, Jacobs RA. Learning optimal integration of arbitrary features in a perceptual discrimination task. Journal of Vision. 2008; 8(3):1–16. [PubMed: 18318629]

Murray RF, Bennett PJ, Sekuler AB. Optimal methods for calculating classification images: weighted sums. Journal of Vision. 2002; 2:79–104. [PubMed: 12678598]

Murray RF, Bennett PJ, Sekuler AB. Classification images predict absolute efficiency. Journal of Vision. 2005; 5:139–149. [PubMed: 15831074]

Nagai M, Bennett PJ, Sekuler AB. Spatiotemporal templates for detecting orientation-defined targets. Journal of Vision. 2007; 7:11. [PubMed: 17685818]

Nandy AS, Tjan BS. The nature of letter crowding as revealed by first- and second-order classification images. Journal of Vision. 2007; 7:1–26. [PubMed: 18217820]

Neri P. Estimation of nonlinear psychophysical kernels. Journal of Vision. 2004; 4:82–91. [PubMed: 15005649]

Neri P, Heeger DJ. Spatiotemporal mechanisms for detecting and identifying image features in human vision. Nature Neuroscience. 2002; 5:812–816.

Neri P, Levi DM. Receptive versus perceptive fields from the reverse-correlation viewpoint. Vision Res. 2006; 46:2465–2474. [PubMed: 16542700]

Neri P, Levi DM. Temporal dynamics of figure-ground segregation in human vision. J Neurophysiol. 2007; 97:951–957. [PubMed: 17005609]

Neri P, Parker AJ, Blakemore C. Probing the human stereoscopic system with reverse correlation. Nature. 1999; 401:695–698. [PubMed: 10537107]

Nestor A, Tarr MJ. Gender recognition of human faces using color. Psychological Science. 2008; 19 in press.

Nolte LW, Jaarsma D. More on the detection of one of M orthogonal signals. Journal of the Acoustical Society of America. 1967; 41:497–505.

Pelli DG, Burns CW, Farell B, Moore-Page DC. Feature detection and letter identification. Vision Res. 2006; 46:4646–4674. [PubMed: 16808957]

Qian J, Richards VM. Relative weights for frequency glide detection using narrowband noise. Journal of the Acoustical Society of America. 2008; 123:2692–2700. [PubMed: 18529188]

Richards VM. Effects of a limited class of nonlinearities on estimates of relative weights. Journal of the Acoustical Society of America. 2002; 111:1012–1017. [PubMed: 11863157]

Richards VM, Tang Z. Estimates of effective frequency selectivity based on the detection of a tone added to complex maskers. Journal of the Acoustical Society of America. 2006; 119:1574–1584. [PubMed: 16583902]

Richards VM, Tang Z, Kidd G Jr. Informational masking with small set sizes. Journal of the Acoustical Society of America. 2002; 111:1359–1366. [PubMed: 11931313]

Richards VM, Zhu S. Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients. Journal of the Acoustical Society of America. 1994; 95:423–434. [PubMed: 8120253]

Ringach DL, Shapley R. Reverse correlation in neurophysiology. Cognitive Science. 2004; 28:147–166.

Schyns PG, Bonnar L, Gosselin F. Show me the features! Understanding recognition from the use of visual information. Psychological Science. 2002; 13:402–409. [PubMed: 12219805]

Sheskin, DJ. Handbook of Parametric and Nonparametric Statistical Procedures. London: Chapman & Hall; 2000.

Shimozaki S, Kingstone A, Olk B, Stowe R, Eckstein M. Classification images of two right hemisphere patients: a window into the attentional mechanisms of spatial neglect. Brain Research. 2006; 1080:26–52. [PubMed: 16497281]

Shimozaki SS, Chen KY, Abbey CK, Eckstein MP. The temporal dynamics of selective attention of the visual periphery as measured by classification images. Journal of Vision. 2007; 7(10):11–20.

Shimozaki SS, Eckstein MP, Abbey CK. Spatial profiles of local and nonlocal effects upon contrast detection/discrimination from classification images. Journal of Vision. 2005; 5:45–57. [PubMed: 15831066]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Solomon JA. Noise reveals visual mechanisms of detection and discrimination. Journal of Vision. 2002; 2:105–120. [PubMed: 12678599]

Stellmack MA, Lutfi RA. Observer weighting of concurrent binaural information. Journal of the Acoustical Society of America. 1996; 99:579–587. [PubMed: 8568045]

Stellmack MA, Viemeister NF. Observer weighting of monaural level information in a pair of tone pulses. Journal of the Acoustical Society of America. 2000; 107:3382–3393. [PubMed: 10875383]

Stellmack MA, Willihnganz MS, Wightman FL, Lutfi RA. Spectral weights in level discrimination by preschool children: analytic listening conditions. Journal of the Acoustical Society of America. 1997; 101:2811–2821. [PubMed: 9165735]

Swets JA, Tanner JWP, Birdsall TG. Decision processes in perception. Psychological Review. 1961; 68:301–340. [PubMed: 13774292]

Tavassoli A, van der Linde I, Bovik AC, Cormack LK. An efficient technique for revealing visual search strategies with classification images. Perception & Psychophysics. 2007a; 69:103–112. [PubMed: 17515220]

Tavassoli A, van der Linde I, Bovik AC, Cormack LK. Orientation anisotropies in visual search revealed by noise. Journal of Vision. 2007b; 7(11):11–18. [PubMed: 17997653]

Thomas JP, Knoblauch K. Frequency and phase contributions to the detection of temporal luminance modulation. Journal of the Optical Society of America A. 2005; 22:2257–2261.

Tjan BS, Nandy AS. Classification images with uncertainty. Journal of Vision. 2006; 6:387–413. [PubMed: 16889477]

Turner CW, Kwon BJ, Tanaka C, Knapp J, Hubbartt JL, Doherty KA. Frequency-weighting functions for broadband speech as estimated by a correlational method. Journal of the Acoustical Society of America. 1998; 104:1580–1585. [PubMed: 9745741]

van Trees, HL. Detection, estimation, and modulation theory. New York: Wiley; 2001.

Victor JD. Analyzing receptive fields, classification images and functional images: challenges with opportunities for synergy. Nature Neuroscience. 2005; 8:1651–1656.

Wang MF, Irwin RJ, Hautus MJ. Detection-theoretic analysis of same-different judgments for the amplitude discrimination of acoustic sinusoids. Journal of the Acoustical Society of America. 2005; 117:1305–1313. [PubMed: 15807019]

Wickens, T. Elementary Signal Detection Theory. Oxford: Oxford University Press; 2001.

Willihnganz MS, Stellmack MA, Lutfi RA, Wightman FL. Spectral weights in level discrimination by preschool children: synthetic listening conditions. Journal of the Acoustical Society of America. 1997; 101:2803–2810. [PubMed: 9165734]

Zwicker, E. Masking and psychological excitation as consequences of the ear's frequency analysis. In: Plomp, R.; Smoorenburg, GF., editors. Frequency Analysis and Periodicity Detection in Hearing. Sijthoff: Leiden; 1970. p. 376-396.

Appendix. Optimal Decision Rule for m-Alternative Spectral-Shape Identification

In this appendix, we derive a likelihood-ratio optimal decision rule for the spectral-shape identification task described in the main text. As a reminder, this task involves m different stimuli, $S_1, \ldots, S_m$, where the subscript, i, denotes which of the n components in the stimulus has an incremented level, compared to the other (equal-amplitude) $n-1$ components. For our spectral-shape identification experiment: $n = m = 5$. Denoting the increment as Δ, and the relative levels of the components in stimulus S_i as Δ_i, we have, for example: $\Delta_1 = [\Delta, 0, \ldots, 0]$, $\Delta_2 = [0, \Delta, 0, \ldots, 0]$, and so forth. On each trial, one stimulus is presented; the task of the participant is to identify which.

In modeling the decision process for this task, we follow other authors (e.g., Durlach, Braida, & Ito, 1986; Green, 1988) in assuming that: 1) each stimulus component evokes a sensory observation, y_i ($i = 1, \ldots, n$); 2) the observations are contaminated by additive Gaussian-distributed noise with zero mean and constant variance; and 3) the noise consists

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of two components, one common to all observations, and the other added independently to the different observations. The latter may result from the randomization of the overall stimulus level across presentations, or from any internal source of noise common to all channels (Dai & Green, 1992).

Denoting as σ_c^2 and σ_I^2 the variances of these two types of noise, the *n*-by-*n* variance-covariance matrix of the observations is:

$$\sum = \begin{bmatrix} \sigma_I^2+\sigma_c^2 & \dots & \sigma_c^2 \\ \vdots & \ddots & \\ \sigma_c^2 & & \sigma_I^2+\sigma_c^2 \end{bmatrix} \tag{A1}$$

The conditional probability density function of the observations, X, given stimulus S_i can then be written as:

$$\begin{aligned} f(Y|S_i) &= \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} e^{-\frac{1}{2}(Y-E[Y])^T\Sigma^{-1}(Y-E[Y])} \\ &= \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} e^{-\frac{1}{2}(Y-\Delta_i)^T\Sigma^{-1}(Y-\Delta_i)} \end{aligned} \tag{A2}$$

where Σ^{-1} and $|\Sigma|$ denotes the inverse and determinant of the variance-covariance matrix, respectively, and the superscript, T, indicates transposition.

In general, the likelihood-ratio optimal decision rule consists in selecting the most likely stimulus alternative given the observations. In the setup considered here, an optimal decision rule is to choose response R_k such that

$$f(Y|S_k) = \max\{f(Y|S_i)\}, \quad i=1,\dots,m \tag{A3}$$

Because $f(Y \mid S_i)$ is related monotonically to the argument of the exponential in Equation A2, an equivalent decision rule uses:

$$Z_i = -(Y-\Delta_i)^T \sum^{-1} (Y-\Delta_i) \tag{A4}$$

as the decision variable. The rule then consists in choosing R_k such that

$$Z_k = max\{Z_i\}, \quad i=1,\dots,m \tag{A5}$$

When, as is typically the case in auditory spectral-shape discrimination experiments (where the overall stimulus level is roved over a wide range relative to threshold), the common variance is much larger than the independent variance of the total noise (i.e., $\sigma_c^2 \gg \sigma_I^2$), the decision variable simplifies to:

$$Z_i = y_i - \frac{\sum_{j\neq i}^{n} y_j}{n-1}, \tag{A6}$$

which is simply the difference between observation y_i and the mean of the other $n-1$ observations. The decision rule then consists in selecting response R_k such that the difference between the observation evoked by the kth stimulus component and the mean of the observations evoked by the other components is largest. This decision rule was used in the Monte Carlo simulations described in the main text.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

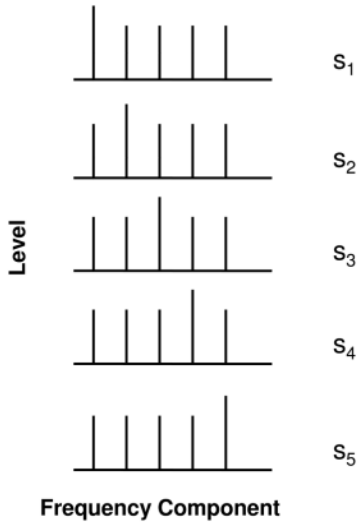

Figure 1.
Spectra of the five stimuli used in the spectral-shape identification experiment. Each spectrum consists of five components. The level of one component is increased by DL (dB) relative to that of the other components. The five resulting stimuli are denoted S_1 to S_5, with the digit indicating the rank of the incremented component.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

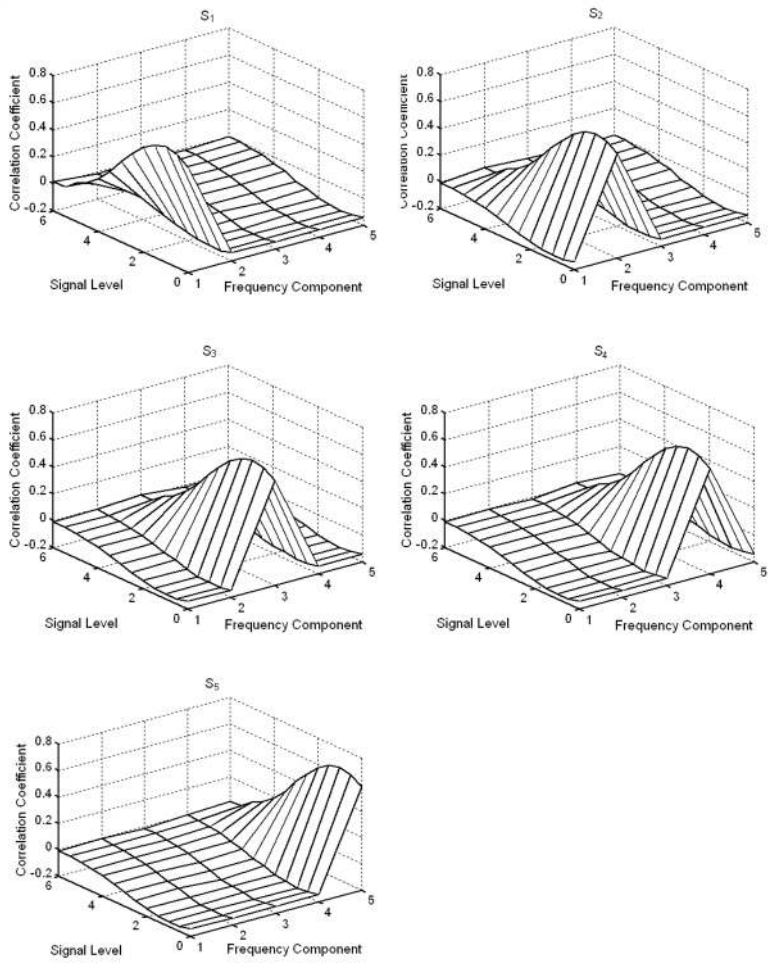

Figure 2.

Estimated weighting patterns, or identification templates, of a simulated optimal observer in an *m*-alternative identification task involving the stimuli shown in Figure 1. The weight estimates are given by point-biserial correlation coefficients between the series of perturbations applied to the levels of the stimulus components, and the corresponding series of response scores (1 for correct, 0 for incorrect), across all trials on which a given stimulus (indicated above each panel) was presented. The coefficients are plotted as a function of the stimulus component number (from 1 to 5), and of the signal strength; the latter was controlled by the increment (DL) applied to the level of the target stimulus component, and varied from 0 to 6 dB.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

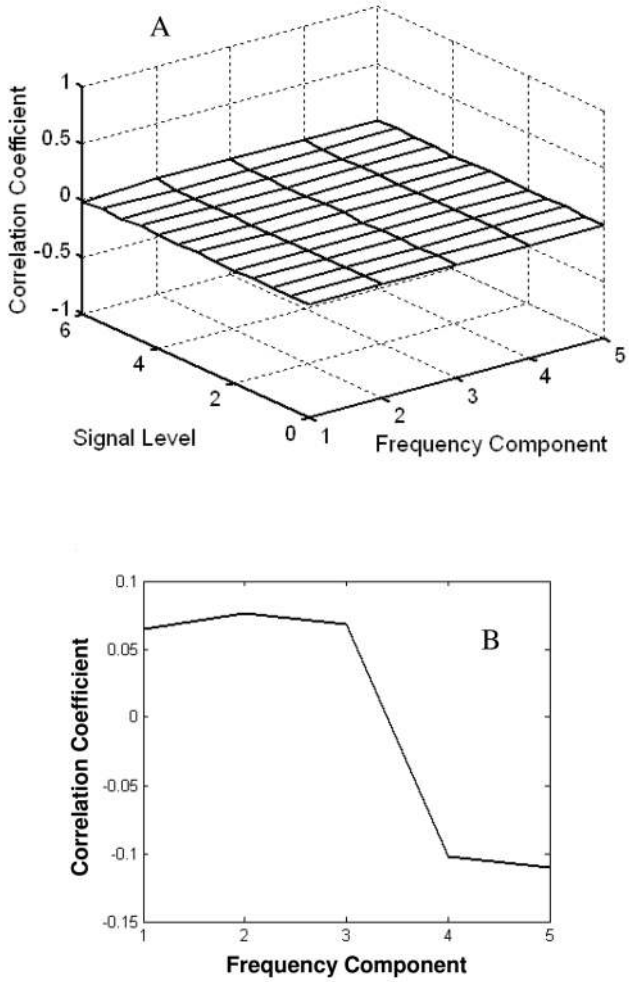

Figure 3.

Influence of indiscriminate across-trial averaging on estimated weighting patterns. **A.** "Overall" weighting pattern estimated by pooling all trials together in the simulated spectral-shape identification experiment involving the five stimuli illustrated in Figure 1. **B.** Weighting pattern obtained by pooling data across all trials in a simulated spectral-shape identification experiment involving only the first three of the five stimuli shown in Figure 1, i.e., S_1, S_2, and S_3. In these simulations, the signal strength (controlled by the level increment, DL, applied to the target stimulus component) was set to 0 dB.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

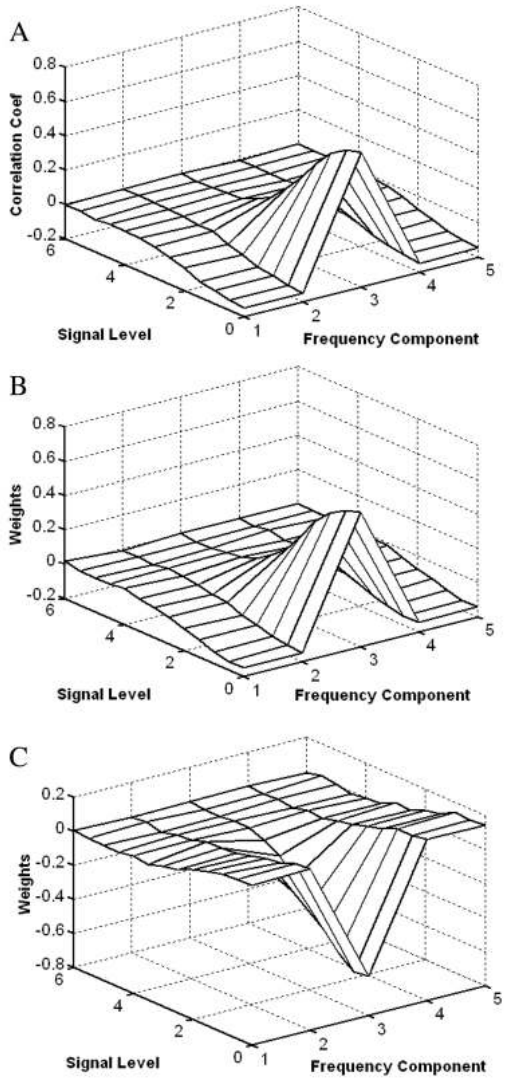

Figure 4.

Example response-based weighting patterns and "classification images". **A.** Response-based weighting patterns based on point-biserial correlation coefficients calculated across all trials between the random perturbations applied to the stimulus-component levels and a binary variable, which was equal to 1 when the response was "S_3", and to 0 otherwise. **B.** Classification image obtained by averaging stimulus–levels perturbations across all trials on which the response was "S_3", regardless of which stimulus alternative was actually presented. To facilitate comparison with the panel S_3 in Figure 2, the weights were scaled by a constant factor so that, for a signal strength (DL) of 0 dB, the weight of the middle component was equal in the two figures. **C.** Classification image obtained by averaging the random perturbations applied to stimulus-component levels across all trials on which the response was different from "S_3", regardless of which stimulus alternative was actually presented. These weights were scaled by the same factor as in panel B.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

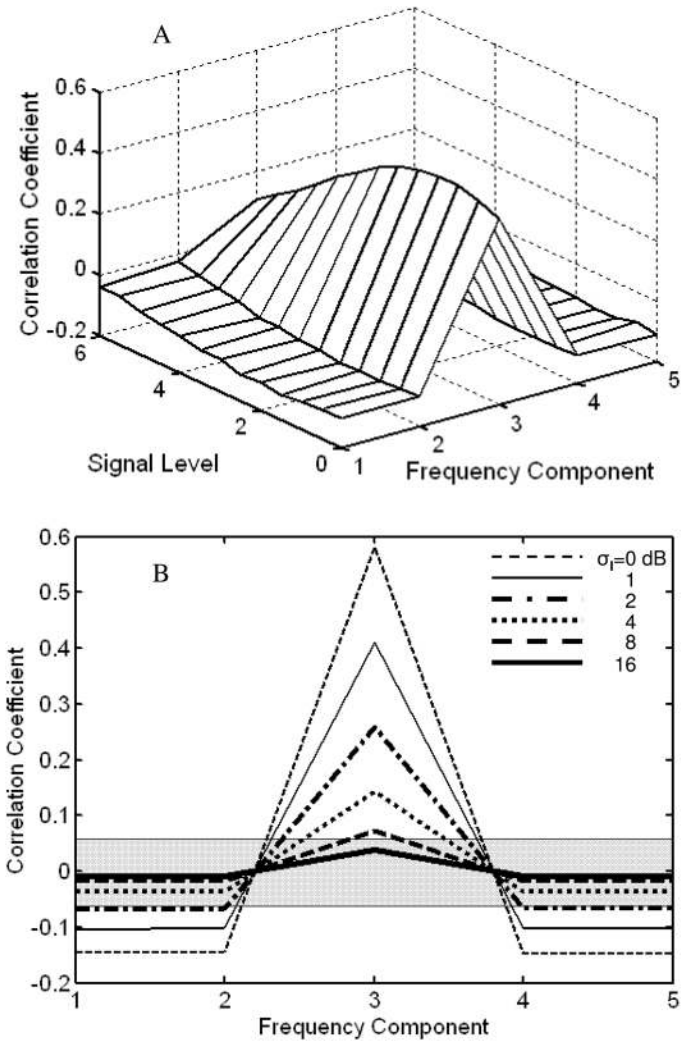

Figure 5.

Influence of internal noise on measured weighting patterns. **A.** Example 3-D weighting pattern obtained in simulations involving internal noise with a standard deviation equal to that of the external perturbations (1 dB). This 3-D weighting pattern is for stimulus S_3; it can be compared with that shown in the panel labeled "S_3" in Figure 2, which was obtained with no internal noise. **B.** Changes in weighting patterns with increasing amounts of internal noise. The different lines indicate weighting patterns obtained in simulations involving internal noise with a standard deviation of 0, 1, 2, 4, 8, or 16 dB, as indicated by the legend. The shaded area around zero correlation represents the 95% confidence interval around a correlation coefficient of zero based on 2000 trials. Only the weighting patterns measured using a null signal strength (DL = 0 dB) is shown here. Note that in this and subsequent figures, only the weighting pattern for stimulus S_3 (i.e., the stimulus with the middle component increased) is shown; the weighting patterns for other stimuli can be obtained from the one shown here by rearranging the frequency-component ranks so that the largest weights correspond to the target component.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

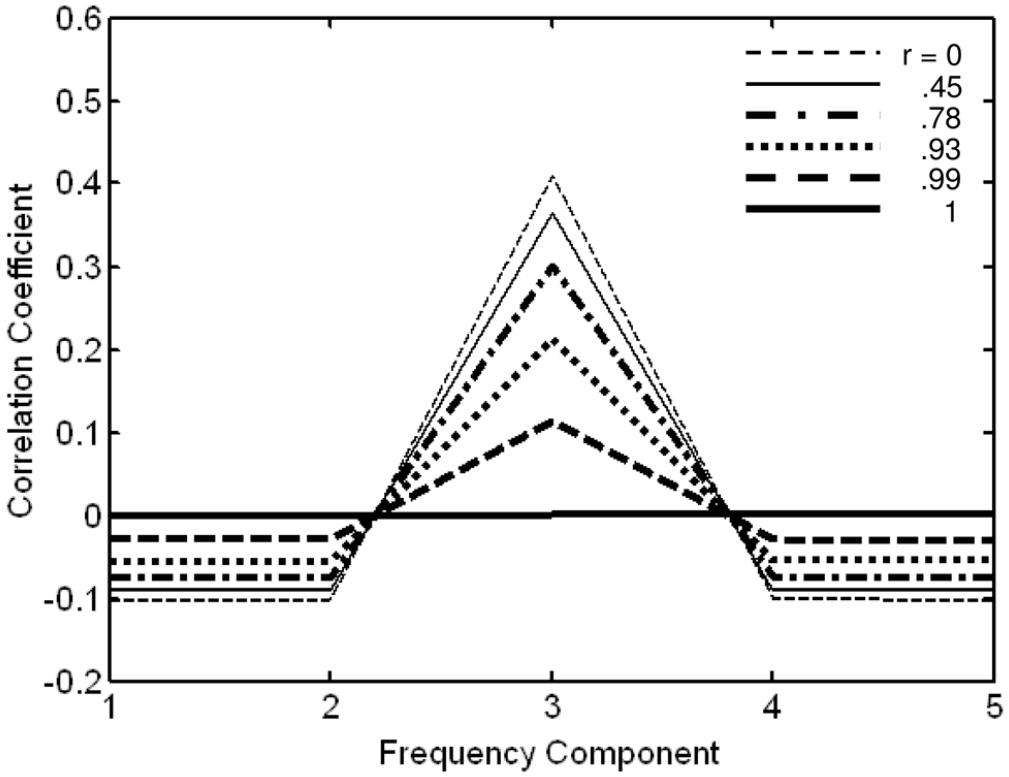

Figure 6.

Influence of correlations between the sensory observations on measured weighting functions. The different line types indicate different degrees of correlation between the target and non-target components: 0, 0.45, 0.78, 0.93, 0.99, and 1. The shaded area around zero correlation represents the 95% confidence interval around a correlation coefficient of zero based on 2000 trials.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

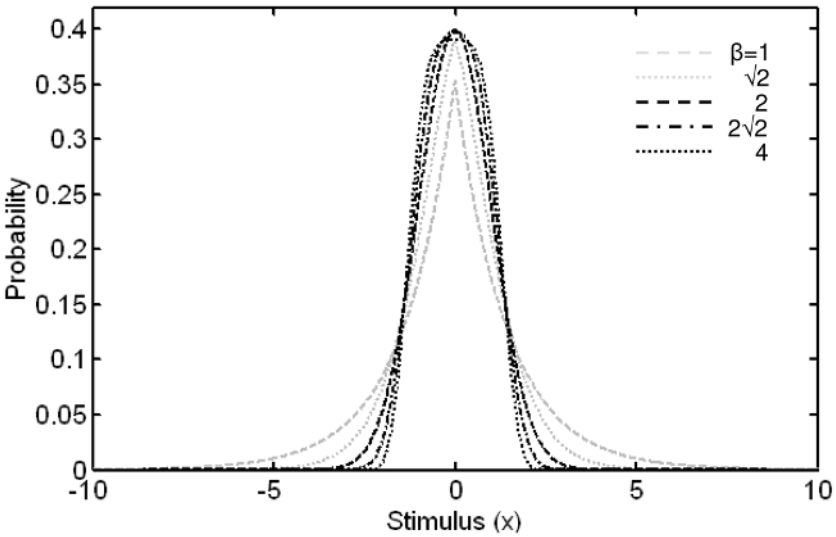

Figure 7.

Probability distributions of stimulus perturbations used in the simulations. These functions were produced by varying the exponent, β, in the function: $p(x)=\kappa\, e^{-|\frac{x}{\sqrt{2}}|^{\beta}}$. This exponent determines the kurtosis of the distribution. The curves are for $\beta=1$ (Laplace, grey-dashed line), $\sqrt{2}$, 2 (Gaussian, dark-dashed line), $2\sqrt{2}$, and 4 (dark-dotted line), from leptokurtic to platykurtic.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

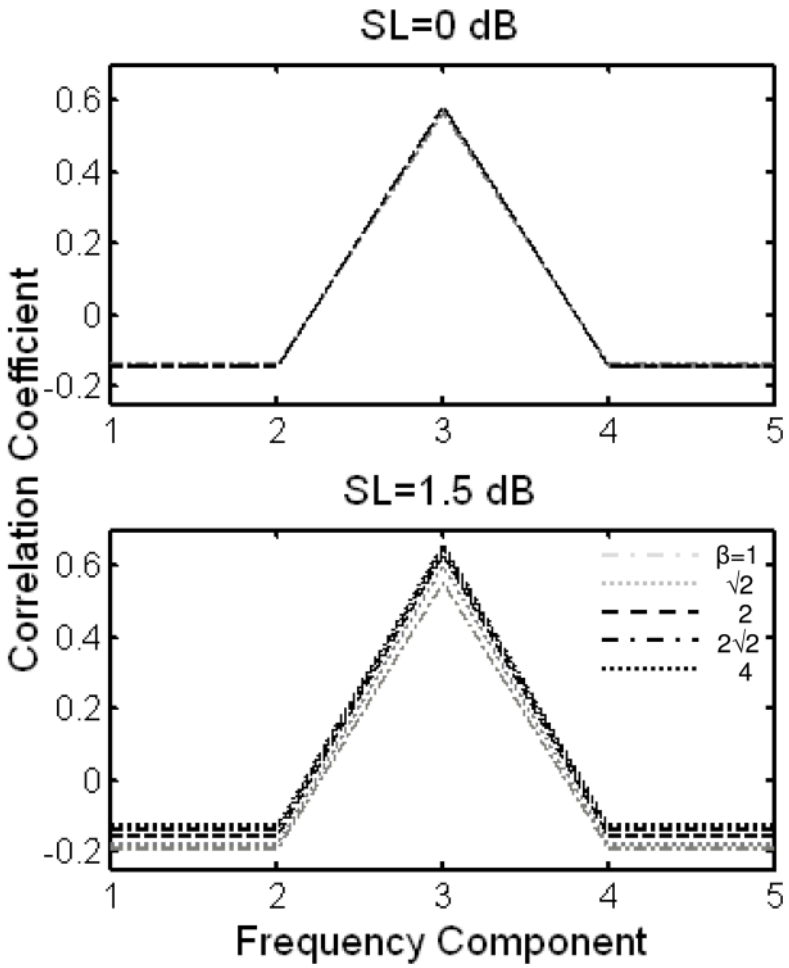

Figure 8.

Measured weighting functions in simulations using the distributions of perturbations shown in Figure 7. **A**. Results for a signal strength, DL, of 0 dB. **B**. Results for a signal strength, DL, of 1.5 dB. The line-style code is the same as in the previous figure.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

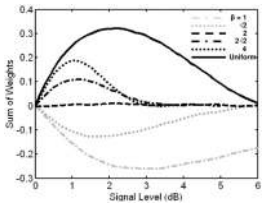

Figure 9.

Influence of the distribution of perturbations on the sum of measured weights as a function of signal strength (controlled by the size of the increment applied to the level of the target stimulus component). Except for the solid line, which corresponds to a uniform distribution, the line-style code is the same as in the previous figure.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

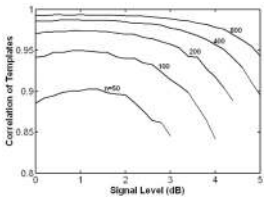

Figure 10.

Mean Pearson correlation coefficients between optimal templates, and templates estimated based on Monte Carlo simulations involving different numbers of trials, ranging from 50 to 800. The correlation coefficients are shown as a function of signal strength (DL), with the number of simulated trials as the parameter; the latter is indicated next to each curve. Each point in this figure represents a mean across 1,000 stochastic simulations.